ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

u b
UNIVERSITÄT BERN

Asking Sensitive Questions in Online Surveys

An Experimental Comparison of the Randomized Response Technique and the Crosswise Model

Andreas Diekmann (ETH Zurich)
Marc Höglinger (ETH Zurich)
Ben Jann (University of Bern)

Kolloquium IMU - Consumer Behavior, Bern, March 22, 2012

Outline

- Misreporting of sensitive issues in (online) surveys
- Some indirect approaches to elicit truthful answers
  - The Randomized Response Technique (RRT)
  - The Crosswise Model (CM): A new alternative to RRT
- Experimental comparison of the different approaches: an online survey on student cheating and plagiarism
- Conclusions

Misreporting of sensitive issues – a pervasive problem

- Respondents might not tell the truth …

Misreporting of sensitive issues – a pervasive problem

- Survey respondents might not tell the truth if asked questions on sensitive issues such as norm violations or deviant behavior. This leads to distorted results.
- Considerable share of 'liars' (respondents with a false negative response) in surveys that use direct questioning (estimates from validation studies):
  - Penal conviction: 42.5% (F2F, Wolter & Preisendörfer 2011)
  - Welfare and unemployment benefit fraud: 75% (F2F, van der Heijden et al. 2000)
  - Driving under influence: 54% (P&P, Locander et al. 1976)
  - Bankruptcy: 32% (Ibid.)

Misreporting of sensitive issues – a problem also in online surveys

- Online surveys offer more anonymity and privacy than interviewer-administered surveys.
- Decrease in the misreporting of sensitive information in online mode compared to CATI (Kreuter, Presser and Tourangeau 2008).
- However, a substantial amount of misreporting remained (Ibd.).
  - falsely denying of poor grade point average:
    83% CATI vs. 62% online mode
  - falsely denying having received an unsatisfactory grade:
    33% CATI vs. 20% online mode

Some Traditional Measurement Techniques

Asking the Embarrassing Question

By Allen H. Barton

University of Chicago

The pollster's greatest ingenuity has been devoted to finding ways to ask embarrassing questions in non-embarrassing ways. We give here examples of a number of these techniques, as applied to the question, "Did you kill your wife?"

1. The Casual Approach:
"Do you happen to have murdered your wife?"

(POQ 22/1958: 67-68)

Some Traditional Measurement Techniques

2. The Numbered Card:

Would you please read off the number on this card which corresponds to what became of your wife?" (HAND CARD TO RESPONDENT)

1. Natural death
2. I killed her
3. Other (What?)

(GET CARD BACK FROM RESPONDENT BEFORE PROCEEDING!)

3. The Everybody Approach:

"As you know, many people have been killing their wives these days. Do you happened to have killed yours?"

. . .

8. Putting the question at the end of the interview.

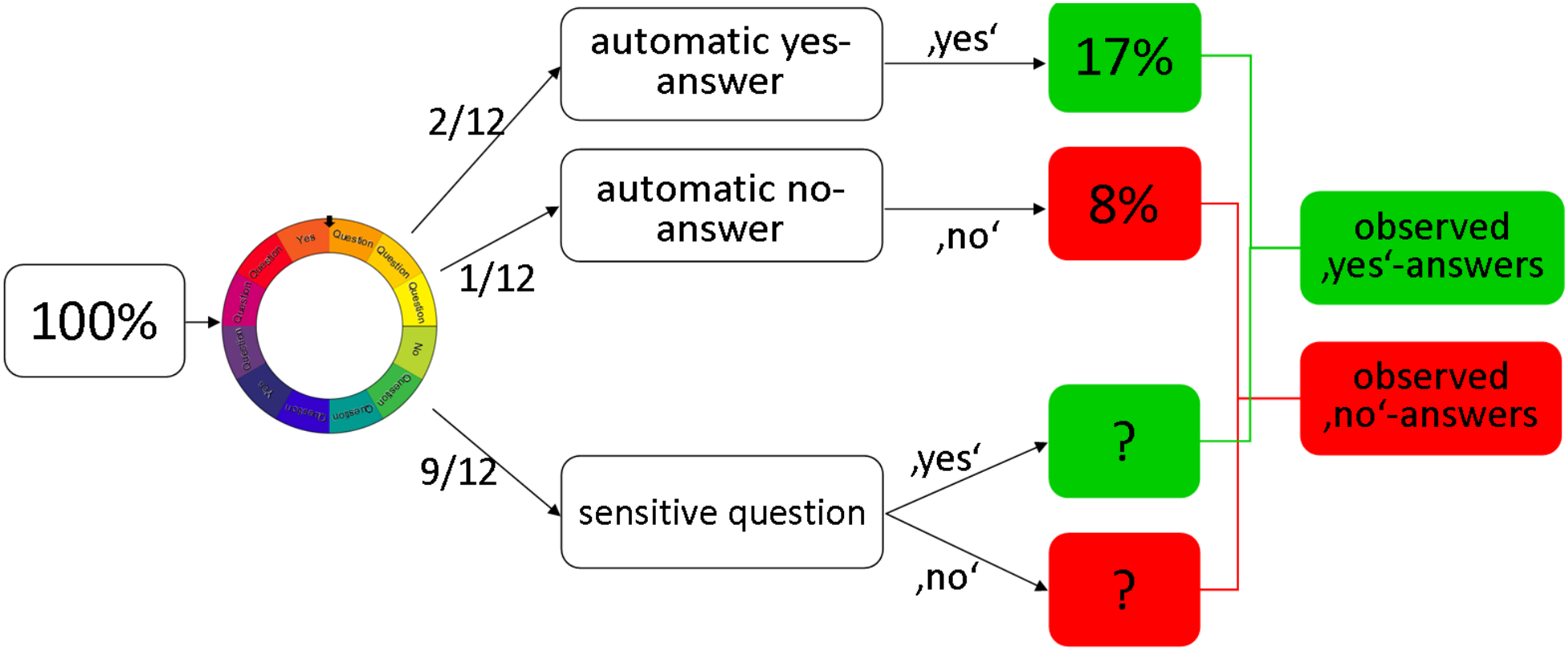
The Randomized Response Technique (RRT)

(Warner 1965; Fox and Tracy 1986)

- Main principle: privacy protection through randomization (i.e. add random noise to the answers)
- A randomizing device, the outcome of which is only known to the respondent, decides whether…
  - the sensitive question has to be answered
  - an automatic ‘yes’ or ‘no’ has to be given or a surrogate question has to be answered
- Since only the respondent knows the outcome of the randomization device, a ‘yes’ cannot be interpreted as an admission of guilt.
- However, with knowledge of the properties of the randomizing device, a prevalence estimate for the sensitive question can be derived.

example

RRT example (forced response design)



- Prevalence estimate (π):

$$Pr(observed\ yes) = Pr(sensitive\ question) * \pi + Pr(automatic\ yes)$$

$$\pi = \frac{Pr(observed\ yes) - P(automatic\ yes)}{Pr(sensitive\ question)}$$

example

The Crosswise Model (CM): a new alternative to RRT

(Yu, Tian, and Tang 2008)

- Simple idea: Ask a sensitive question and a non-sensitive question and let the respondent indicate whether …
  - the answers to the questions are the same (both 'yes' or both 'no')
  - the answers to the questions are different (one 'yes', the other 'no')

		non-sensitive question	
		No	Yes
sensitive question	No	same	different
	Yes	different	same

- Note: Questions must be uncorrelated and the probability of answering 'yes' to the non-sensitive question must be unequal 0.5.

example

The Crosswise Model (CM): a new alternative to RRT

(Yu, Tian, and Tang 2008)

- Prevalence estimate (π):

$$Pr(same) = (1 - \pi) * (1 - \Pr(nonsensitive\ yes)) + \pi * \Pr(nonsensitive\ yes)$$

$$\pi = \frac{Pr(same) + Pr(nonsensitive\ yes) - 1}{2 * Pr(nonsensitive\ yes) - 1}$$

- Note: CM is formally identical to Warner's original RRT model.

example

The Crosswise Model: Let's practice

- Two questions:

  1. Is your mother's birthday in January or February?
     (You can also think of someone else. It doesn't matter.)

  2. Did you ever falsify your data or results?
     (e.g. edit data points or delete observations so that hypothesis is confirmed, falsify entire dataset, invent or manipulate reported results)

- Compare your answers: Are they the same or different?

  - Write "**A**" if they are the same (both Yes or both No)

  - Write "**B**" if they are different (one Yes, the other No)

Generalized estimator for RRT and CM

- Let

  Y_i response ($Y_i = 1$ if "yes" in RRT or "A" in CM, else $Y_i = 0$)

  λ_i probability of $Y_i = 1$

  π_i (unknown) prevalence of sensitive item

  p_i^{w} probability of being directed to the negated question in Warner's RRT (or prevalence of nonsensitive item in CM)

  p_i^{yes} overall probability of surrogate "yes"

  p_i^{no} overall probability of surrogate "no"

- Then

$$\lambda_i = (1 - p_i^{\text{yes}} - p_i^{\text{no}})p_i^{\text{w}}\pi_i + (1 - p_i^{\text{yes}} - p_i^{\text{no}})(1 - p_i^{\text{w}})(1 - \pi_i) + p_i^{\text{yes}}$$

and hence

$$\pi_i = \frac{\lambda_i - (1 - p_i^{\text{yes}} - p_i^{\text{no}})(1 - p_i^{\text{w}}) - p_i^{\text{yes}}}{(2p_i^{\text{w}} - 1)(1 - p_i^{\text{yes}} - p_i^{\text{no}})}$$

Two Stata commands (available from the SSC Archive)

`rrreg` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *regress_options*
`pwarner(`#|*varname*`)` `pyes(`#|*varname*`)` `pno(`#|*varname*`)`]

- Assumes $\pi_i = X_i'\beta$ and estimates β using least squares with transformed response

$$\tilde{Y}_i = \frac{Y_i - (1 - p_i^{\text{yes}} - p_i^{\text{no}})(1 - p_i^{\text{w}}) - p_i^{\text{yes}}}{(2p_i^{\text{w}} - 1)(1 - p_i^{\text{yes}} - p_i^{\text{no}})}$$

Two Stata commands (available from the SSC Archive)

`rrlogit` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *logit_options*
`pwarner(#|varname) pyes(#|varname) pno(#|varname)`]

- Assumes $\pi_i = e^{X_i'\beta}/(1+e^{X_i'\beta})$ and estimates β using maximum likelihood with

$$\ln L = \sum_{i=1}^{n} \left\{ Y_i \ln(R_i) + (1-Y_i)\ln(S_i) - \ln(1+e^{X_i'\beta}) \right\}$$

where

$$R_i = c_i + q_i e^{X_i'\beta} \qquad c_i = (1 - p_i^{\text{yes}} - p_i^{\text{no}})(1 - p_i^{\text{w}}) + p_i^{\text{yes}}$$
$$S_i = (1-c_i) + (1-q_i)e^{X_i'\beta} \qquad q_i = (1 - p_i^{\text{yes}} - p_i^{\text{no}})p_i^{\text{w}} + p_i^{\text{yes}}$$

Performance of RRT in online mode

- RRT does not seem to work well in in online surveys. Results so far showed for the RRT…
  - no difference in or even lower prevalence estimates for socially undesirable behavior compared to direct questioning (Coutts et al. 2011 , Coutts & Jann 2011, Peeters 2006, Snijders & Weesie 2008)
  - unrealistically high prevalence estimates for voting (Holbrook & Krosnick 2010)
  - exception: higher prevalence estimates with the RRT in a survey on adult entertainment desires (de Jong, Pieters and Fox 2010)
- However, RRT implementations so far were often not well suited to online mode.
  - randomizing device not at respondents' immediate reach
  - randomizing device not trustworthy

Performance of the Crosswise Model

- The Crosswise Model seems to be a promising alternative
  - higher prevalence estimates than with direct questioning in a p&p survey on plagiarism (Jann, Jerke, Krumpal 2011)
  - however, no empirical application in online mode so far
- Advantages of the Crosswise Model over RRT
  - easier to understand
  - no need for a randomizing device
  - respondent is not forced into giving a ‘false’ automatic response
  - no obvious self-protective answering strategy (e.g. always tick ‘no’)

Our study: survey on student cheating and plagiarism

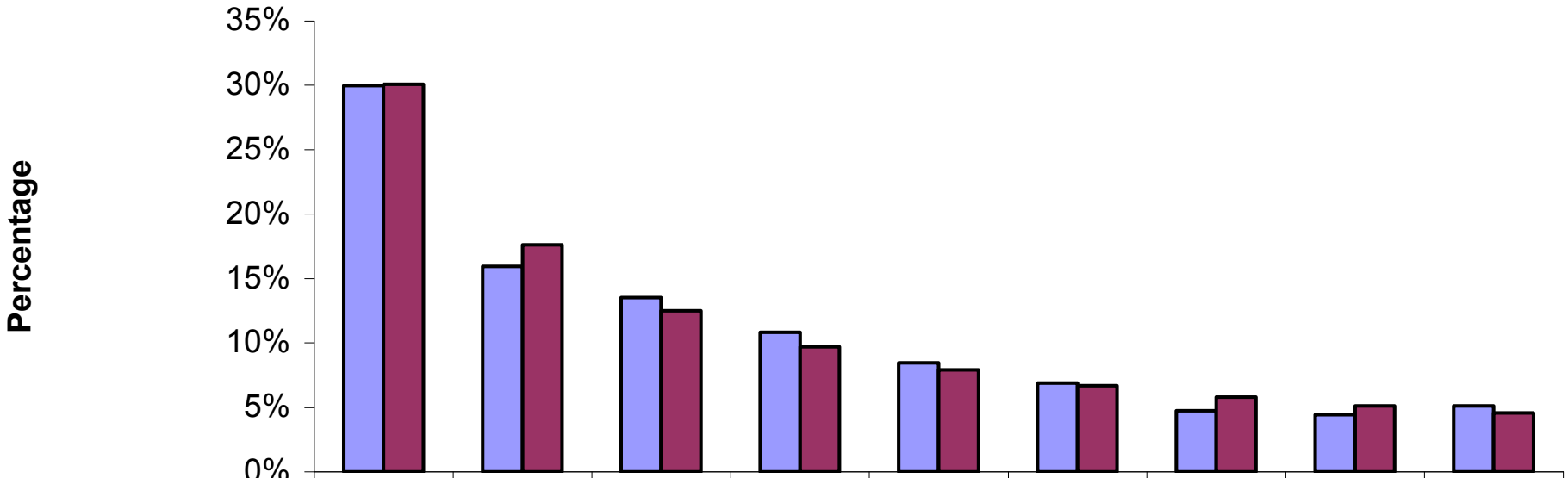
- Web survey among students of the University of Bern and the ETH Zurich in spring 2011
- Response rate 33%, 6'494 completed interviews
- Sensitive questions on
  - copying from other students in exam (copy)
  - using crib notes in exam (notes)
  - taking drugs to enhance exam performance (drugs)
  - partial paper plagiarism (partial)
  - severe paper plagiarism (severe)
- Comparing direct questioning (DQ) to three variants of RRT and two variants of the Crosswise Model (CM)
- Aprox. 1'000 randomly assigned respondents in each experimental condition

Experimental conditions: 6 different implementations of the sensitive questions

- DQ: direct questioning
  >
- RRT wheel: forced response RRT using virtual random wheel
  >
- RRT pick: forced response RRT using ‘Pick a number’ method
  >
- RRT Benford: RRT using Benford distribution and unrel. questions
  >, >>
- CM unr. quest.: Crosswise Model using unrelated questions
  >
- CM pick: Crosswise Model using ‘Pick a number’ method
  >

Benford's Law: Pr(d) = log10(1 + 1/d) with d in {1,2,...,9}

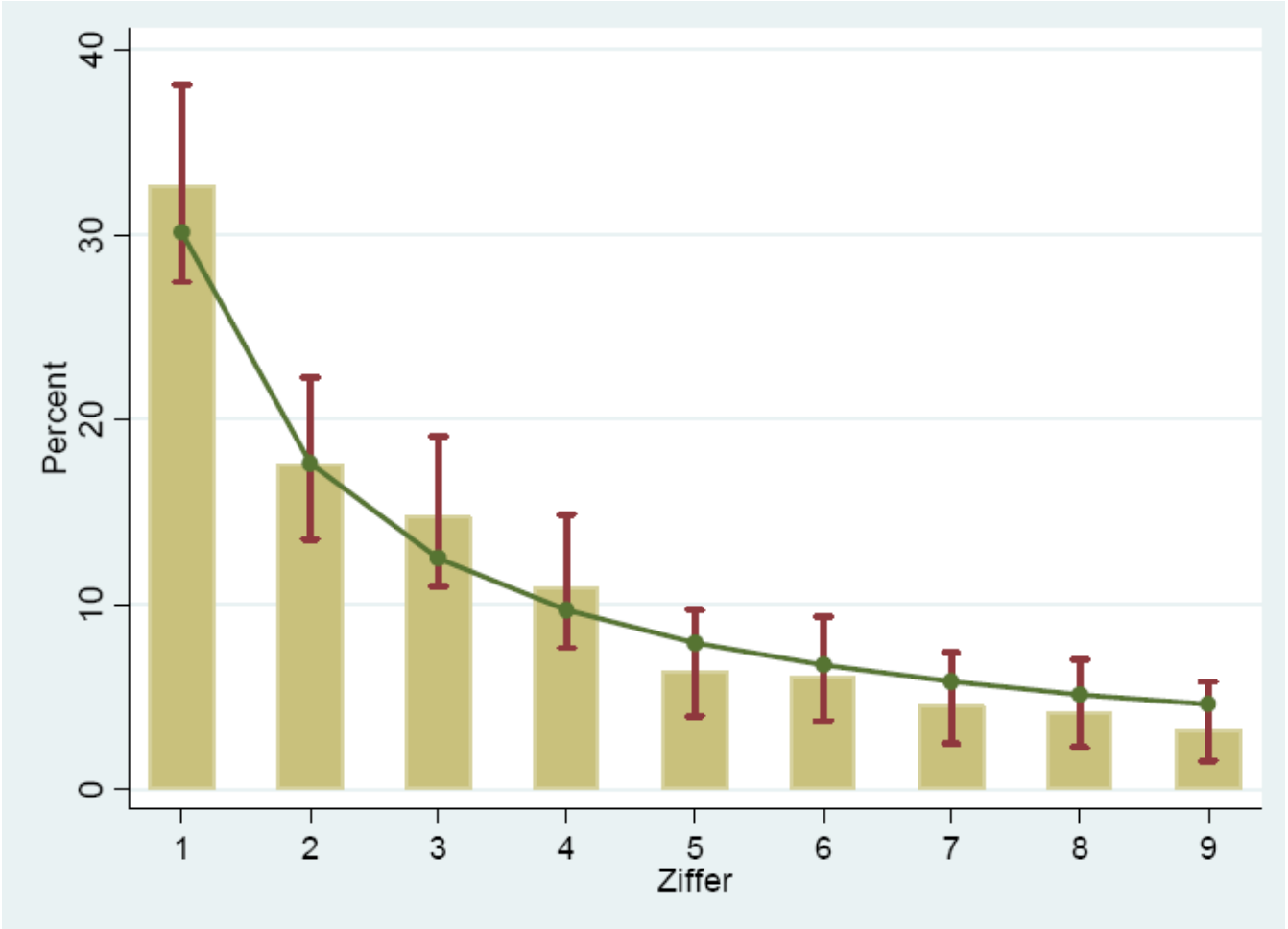
House numbers collected from the telephone directory of Zurich



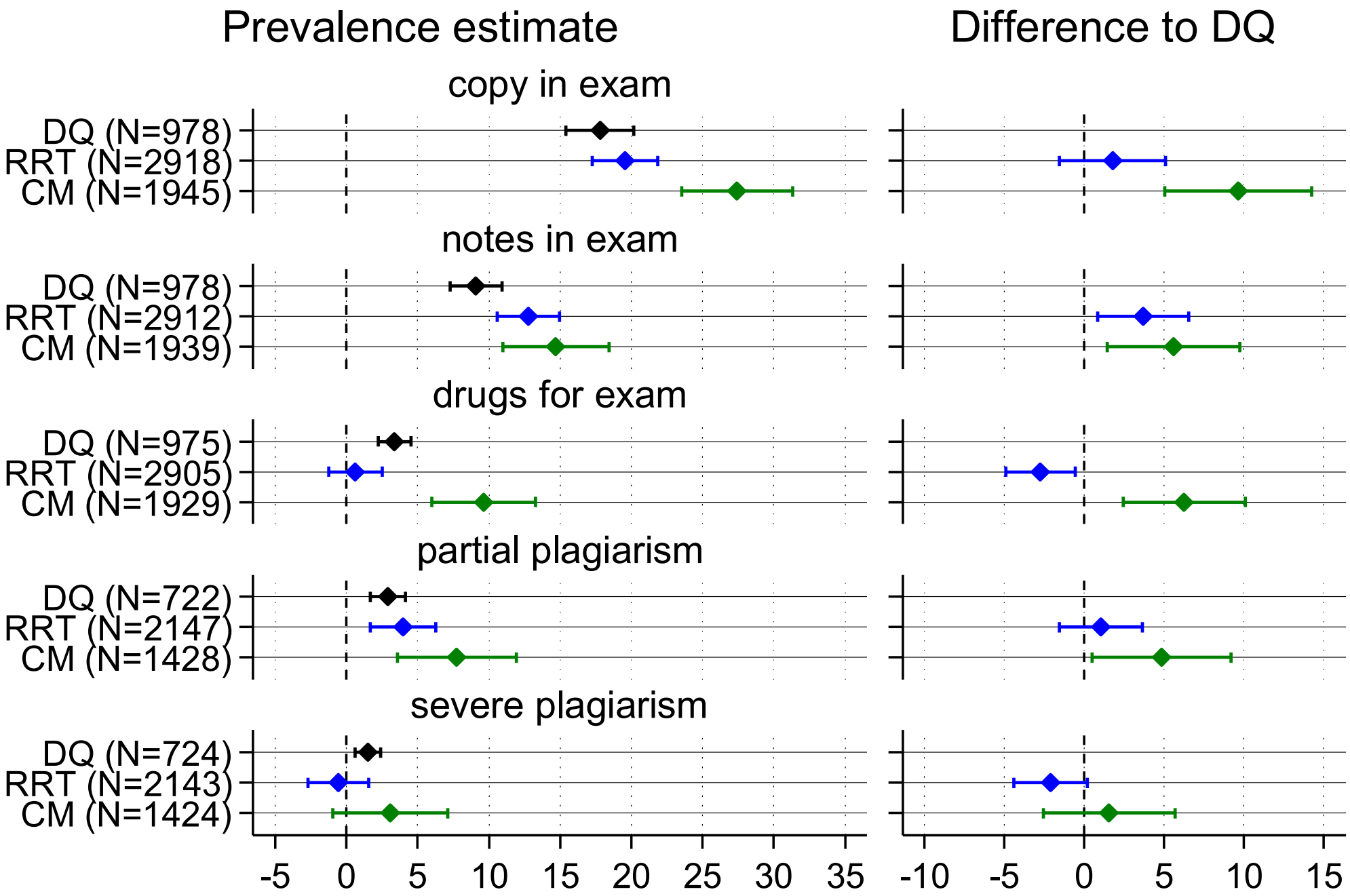
	1	2	3	4	5	6	7	8	9
House number	29,99%	15,96%	13,51%	10,84%	8,46%	6,90%	4,75%	4,45%	5,12%
Benford	30,10%	17,61%	12,49%	9,69%	7,92%	6,69%	5,80%	5,12%	4,58%

Denken Sie an eine Ihnen bekannte Person, von der Sie die Wohnadresse wissen. Bitte geben Sie die **erste Ziffer** der Hausnummer der Adresse dieser Person an.

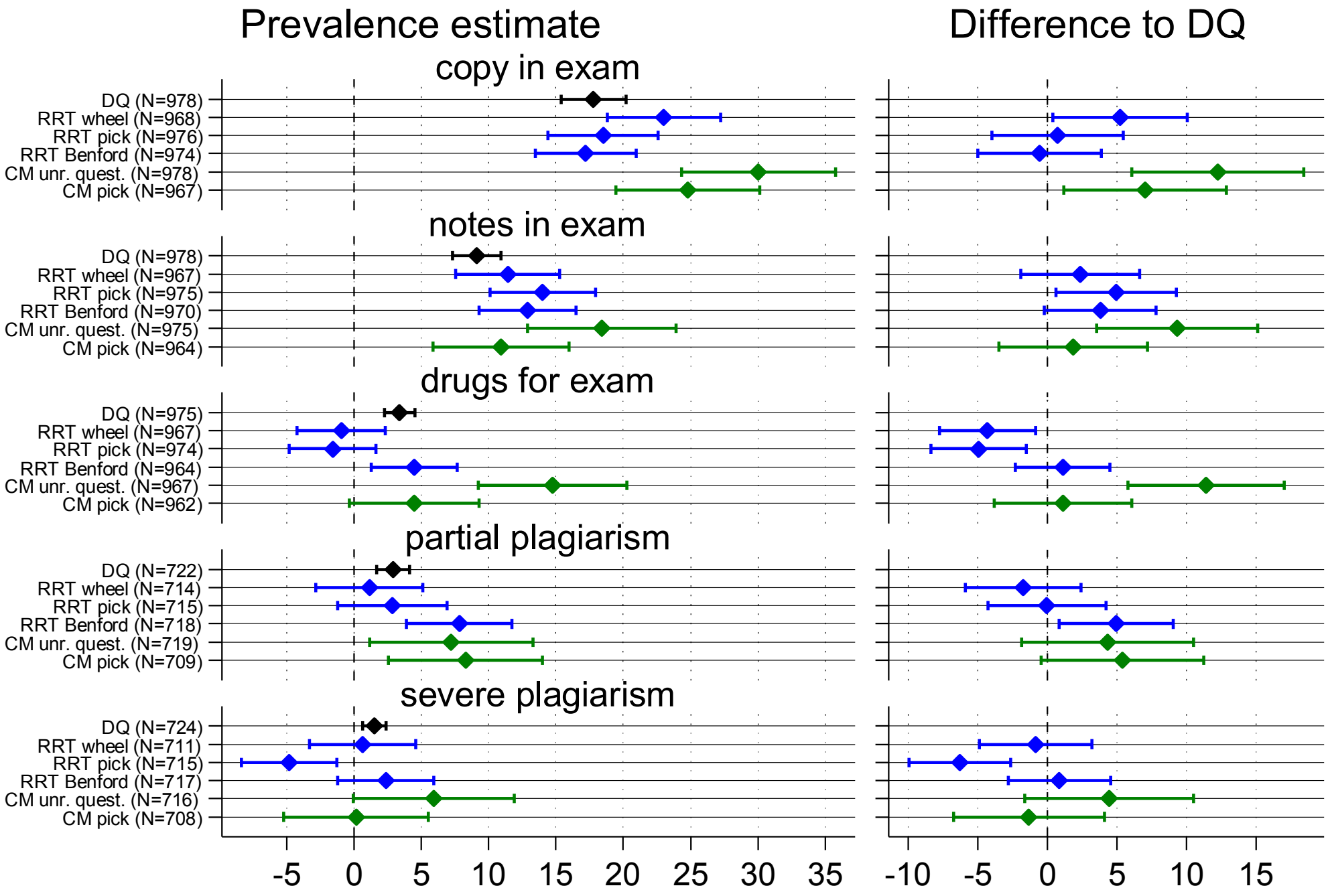
The answers follow Benford‘s Law! (chi2 = 6.2, 8 df, p = 0.62)



Prevalence estimates (with 95%-ci) by technique

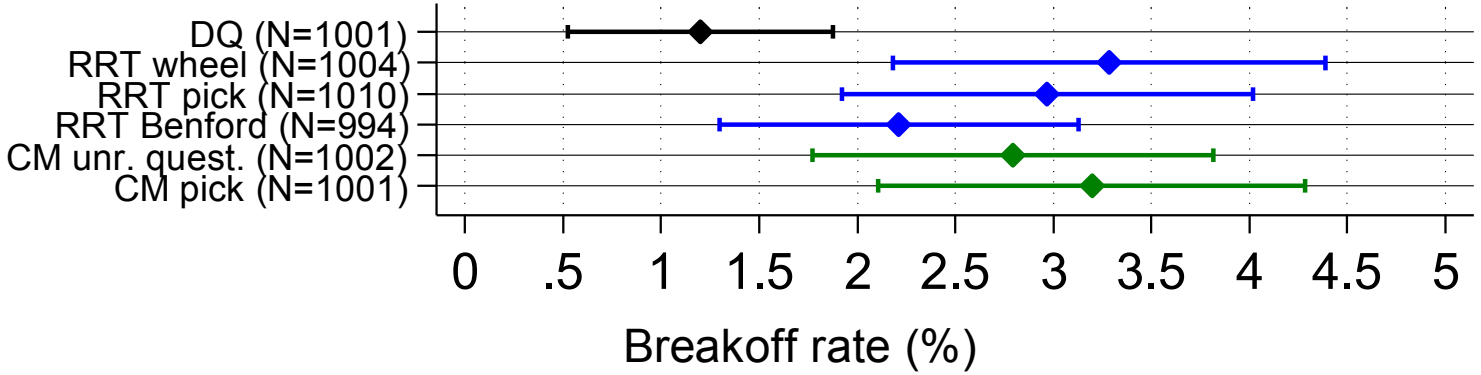


Prevalence estimates (with 95%-ci) by implementation

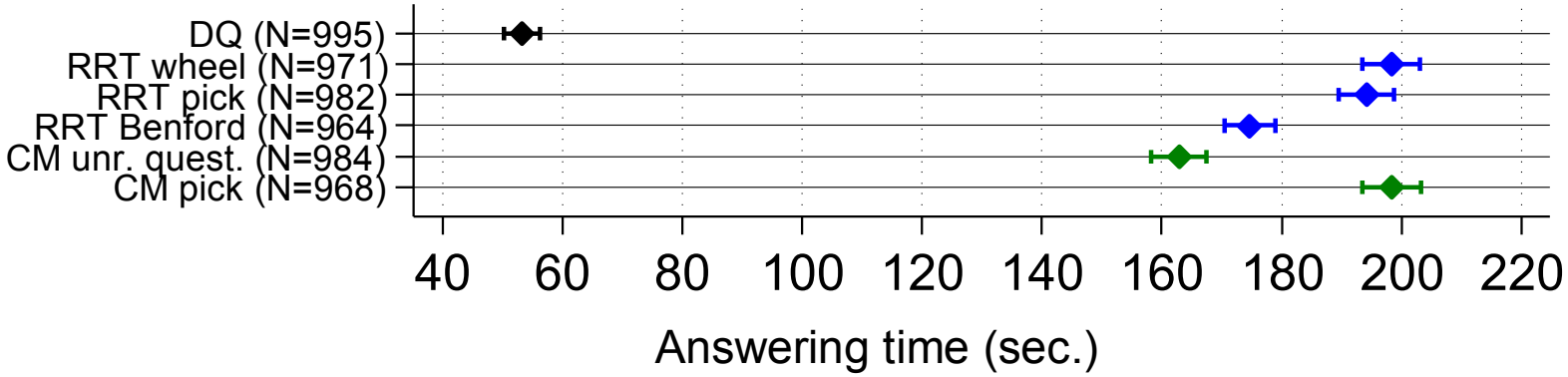


Breakoff rates and response time by implementation

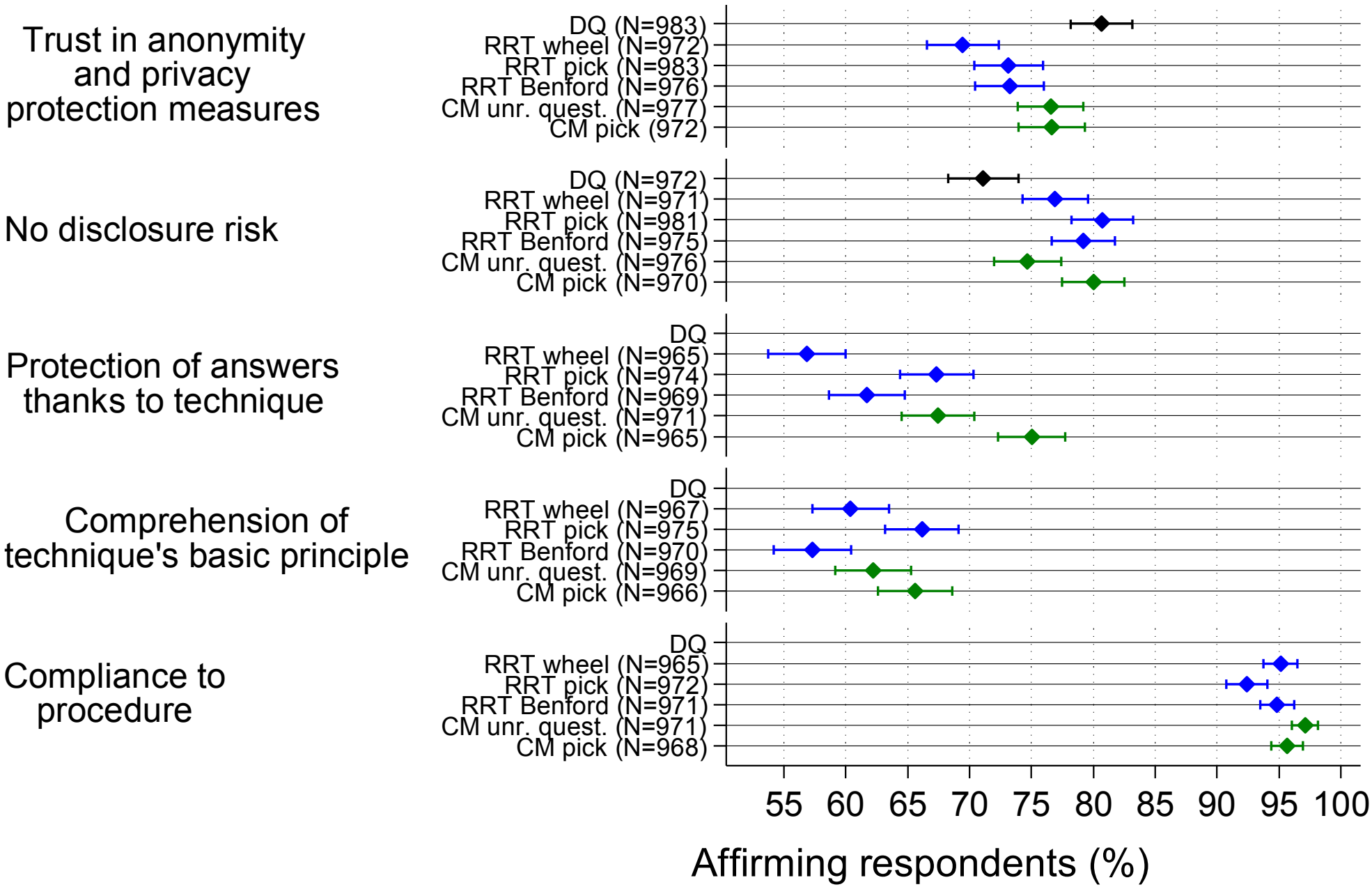
Breakoff after start sensitive items



Time to answer sensitive items



Respondents' evaluation by implementation



Determinants of sensitive behavior

Randomized response logistic regression

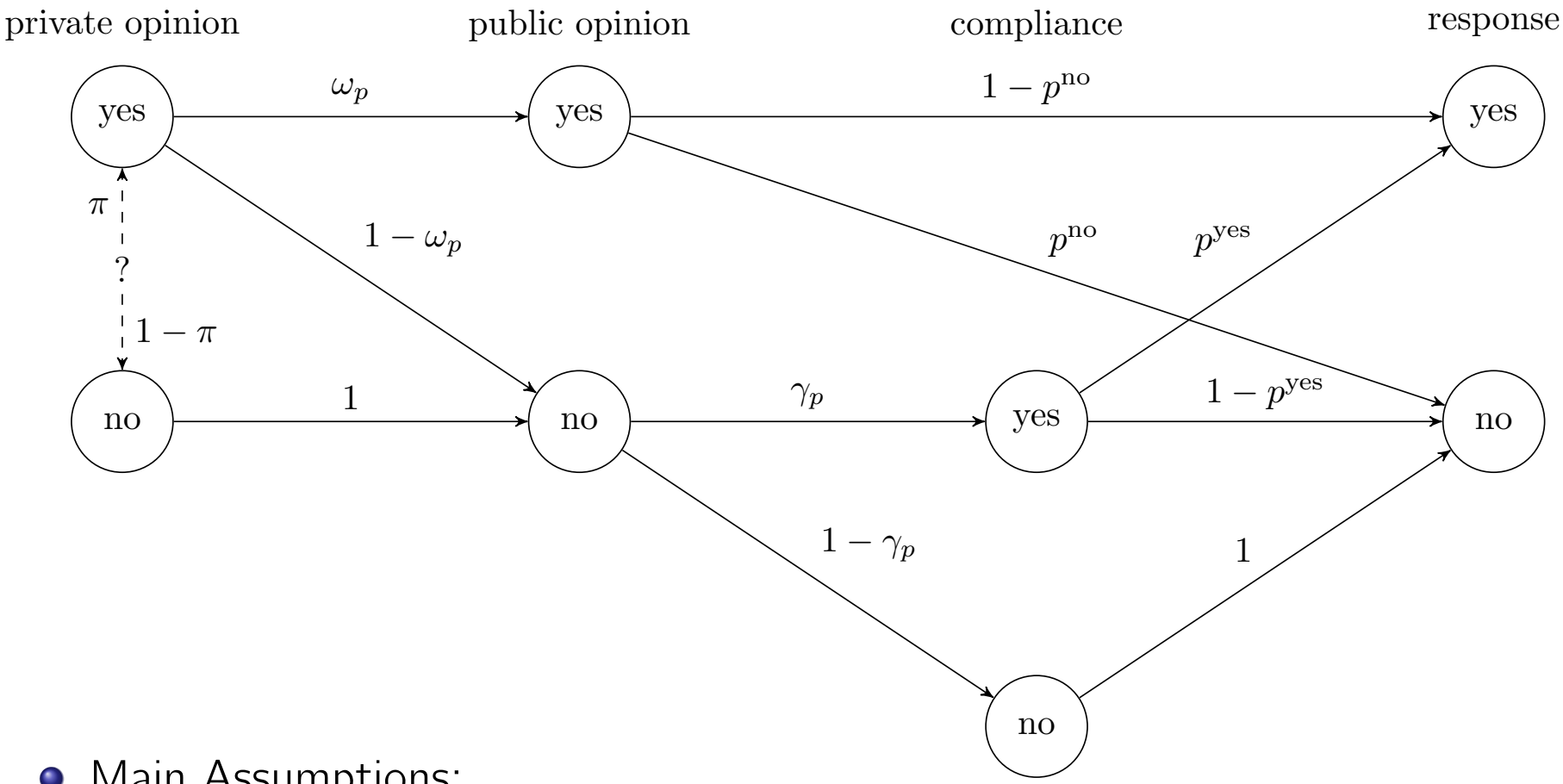
	copy	notes	drugs	partial	severe
ETH (ref. UniBE)	-0.150	0.192	-0.577*	0.386	1.032
	(0.104)	(0.140)	(0.264)	(0.308)	(0.697)
Semester (log)	0.072	0.298	-0.026	-0.120	-0.589
	(0.163)	(0.211)	(0.355)	(0.323)	(0.523)
Nbr. exams/papers (log)	0.596***	0.298*	-0.264	0.393*	0.208
	(0.111)	(0.135)	(0.211)	(0.196)	(0.398)
Perceived risk	-0.014***	-0.024***		-0.008	-0.009
	(0.004)	(0.006)		(0.006)	(0.013)
Risk attitude	0.066**	0.095**	0.184**	0.144	-0.037
	(0.025)	(0.032)	(0.066)	(0.078)	(0.132)
Procrastination	0.202***	0.213**	0.189	0.275	0.001
	(0.049)	(0.066)	(0.128)	(0.152)	(0.320)
Stress at exams	0.104*	0.084	0.462***		
	(0.050)	(0.066)	(0.134)		
Constant	-4.615***	-5.202***	-5.791***	-5.843***	-3.706*
	(0.414)	(0.596)	(0.934)	(1.065)	(1.521)
N	5713	5713	5761	4224	4221

Standard errors in parentheses
Additional controls for experimental conditions
* p<0.05, ** p<0.01, *** p<0.001

Summary

- The Crosswise Model produced significantly higher prevalence estimates and, therefore, clearly outperformed DQ (if we accept the 'more-is-better'-assumption)
  - An exception is the last item (severe plagiarism) with a very low prevalence.
- RRT, on the other hand, does not yield higher estimates than DQ (even lower and sometimes negative estimates).
  - One reason might be the 'self-protective no' bias, which prevents respondents to say 'yes' if instructed to do so by the randomizing device.

A little bit of magic: Cheating detection in RRT



- Main Assumptions:
  - Monotonicity of social desirability: Public opinion is always "no" if private opinion is "no"
  - No provocation: Respondents do not say "yes" if advised to say "no"

A little bit of magic: Cheating detection in RRT

- Assuming that γ and ω do not depend on p^{yes} and p^{no} (which may be justified if variation in p is small) (and that γ does not depend on the private opinion), this leads to the following log likelihood:

$$\ln L = \sum_{i=1}^{n} Y_i \ln(\ell_i) + (1 - Y_i) \ln(1 - \ell_i)$$

with

$$\ell_i = \pi_i \omega (1 - p_i^{no} - \gamma p_i^{yes}) + \gamma p_i^{yes}$$

- If p^{yes} and p^{no} are randomly varied between respondents, then $\pi_i \omega$ and γ are identified.

A little bit of magic: Cheating detection in RRT

	copy	notes	drugs	partial	severe
RRT adjusted	17.9	12.0	16.7	14.3	6.7
	(6.5)	(6.1)	(5.6)	(6.6)	(5.9)
Cheaters	-9.5	-3.6	88.9	54.3	36.1
	(36.1)	(31.9)	(36.9)	(40.1)	(31.8)
N	2855	2855	2849	2105	2104

Standard errors in parentheses

Unadjusted results for comparison:

	copy	notes	drugs	partial	severe
DQ	17.5	8.8	3.4	2.5	1.5
	(1.2)	(0.9)	(0.6)	(0.6)	(0.5)
RRT	19.6	12.7	0.6	4.2	-0.6
	(1.2)	(1.1)	(1.0)	(1.2)	(1.1)
CM	27.2	15.0	9.9	8.2	3.0
	(2.0)	(1.9)	(1.9)	(2.1)	(2.0)

Conclusions

- RRT does not seem to be a good method for online surveys.
  - Although a lot of effort has been put into pretesting and finding good implementations, no convincing evidence could be found that RRT yields more valid estimates than DQ.
  - With RRT 'Benford' performing somewhat better than the other RRT implementations…
- The Crosswise Model is a promising alternative, since it does not suffer from some of the deficiencies of the RRT ("self-protective no" bias, complexity).
- Improvement of RRT estimates is possible by correcting for cheating respondents who do not comply with the instructions. Such estimates, however, have low efficiency.

Substantive conclusions
(based on combined results from CM)

- A substantial proportion of students cheated on an exam (copying: about 25 percent, crib notes: about 15 percent)
- Using drugs to enhance exam performance is not uncommon (10 percent)
- Rates for partial plagiarism (using a passage from someone else's work without providing proper citation) are 8 percent. The prevalence of severe plagiarism (hand in someone else's work) is 3 percent.
- These numbers may not seem excessively high, but we have to keep in mind that they most likely still underestimate the true prevalence.

Thank you!

References I

Coutts, E. und B. Jann. 2011. Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT), *Sociological Methods & Research*, Vol. 40, 196-193.

Coutts, Elisabeth, Benn Jann, Ivar Krumpal, and Anatol-Fiete Näher. 2011. "Plagiarism in Student Papers: Prevalence Estimates Using Special Techniques for Sensitive Questions." *Journal of Economics and Statistics 231:749-760.*

de Jong, Martijn G., Rik Pieters, and Jean-Paul Fox. 2010. "Reducing Social Desirability Bias Through Item Randomized Response: An Application to Measure Underreported Desires." *Journal of Marketing Research (JMR)* 47:14-27.

Edgell, S. E., S. Himmelfarb und K. L. Duchan. 1982. Validity of forced responses in a randomized-response model, *Sociological Methods & Research*, Vol. 11, 89-100.

Fox, J. A. und P. E. Tracy. 1986. *Randomized response: A method for sensitive surveys*, Newbury Park, CA: Sage.

Holbrook, A. L. und J. A. Krosnick. 2010. Measuring Voter Turnout By Using The Randomized Response Technique: Evidence Calling Into Question The Method's Validity, *Public Opin Q*, Vol. 74, 328-343.

Jann, B., J. Jerke und I. Krumpal. 2011. Asking Sensitive Questions Using the Crosswise Model: Some Experimental Results, *Public Opinion Quarterly*.

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and Web Surveys." *Public Opinion Quarterly* 72:847-865.

References II

Lensvelt-Mulders, G. J. L. M. und H. R. Boeije. 2007. Evaluating compliance with a computer assisted randomized response technique: a qualitative study into the origins of lying and cheating, *Computers in Human Behavior*, Vol. 23, 591-608.

Locander, W., S. Sudman und N. Bradburn. 1976. An Investigation of Interview Method, Threat and Response Distortion, *Journal of the American Statistical Association*, Vol. 71, 269-275.

Peeters, C. F.W. 2006. Measuring politically sensitive behavior. Using probability theory in the form of randomized response to estimate prevalence and incidence of misbehavior in the public sphere: a test on integrity violations. Amsterdam: Faculty of Social Sciences, Vrije Universiteit Amsterdam.

Snijders, C., and J. Weesie. 2008. "The online use of randomized response measurement." Paper presented at General Online Research 2008, Hamburg, Germany.

van der Heijden, P. G. M., G. van Gils, J. Bouts und J. J. Hox. 2000. A Comparison of Randomized Response, Computer-Assisted Self-Interview, and Face-to-Face Direct Questioning. Eliciting Sensitive Information in the Context of Welfare and Unemployment Benefit, *Sociological Methods and Research*, Vol. 28, 505–537.

Warner, S. L. 1965. Randomized-response: A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, Vol. 60, 63-69.

Wolter, F., and P. Preisendörfer. 2011. "Asking Sensitive Questions: An Evaluation of the Randomized Response Technique vs. Direct Questioning Using Individual Validation Data." unpublished.

Yu, J.-W., G.-L. Tian, M.-L. Tang (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. Metrika 67(3): 251-263.

Appendix: Items

Wording sensitive items

	Item	Wording (translated from German)
1	copying from fellow students in exam	Have you ever copied from fellow students in exam during your studies?
2	using crib notes in exam	Have you ever illicitly used crib notes in exam during your studies (also notes on mobile phones, calculators or similar)?
3	taking drugs to enhance exam performance	Have you ever used drugs only available on prescription to enhance your exam performance during your studies?
4	partial paper plagiarism	Have you ever deliberately taken a whole passage from another source without marking it as a quote in a submitted paper during your studies?
5	severe paper plagiarism	Have you ever had someone else writing the bulk of a submitted paper or have you handed in someone else's paper as your own during your studies?

Wording evaluation questions

Zum Abschluss interessiert uns noch Ihre Einschätzung dieser Umfrage:

Bitte ganz ehrlich: Wie stark vertrauen Sie unseren Massnahmen zur Anonymität und zum Persönlichkeitsschutz der Teilnehmenden bei dieser Umfrage?

gar nicht	eher nicht	teils, teils	eher stark	sehr stark
○	○	○	○	○

Für wie wahrscheinlich halten Sie es, dass aufgrund dieser Umfrage nachvollzogen werden kann, ob eine bestimmte Teilnehmerin/ein Teilnehmer dieser Umfrage eines der erfragten heiklen Verhalten (Abschreiben, Spickzettel, Plagiate etc.) **begangen hat?**

unmöglich	sehr un-wahrscheinlich	eher un-wahrscheinlich	eher wahrscheinlich	sehr wahrscheinlich
○	○	○	○	○

Wording evaluation questions

Um Ihren Persönlichkeitsschutz bedingungslos sicherzustellen, haben wir bei einzelnen heiklen Fragen eine spezielle Befragungsmethode benutzt:

Wie umständlich war für Sie das Befolgen dieser speziellen Befragungsmethode?

sehr umständlich	eher umständlich	teils, teils	eher nicht umständlich	überhaupt nicht umständlich
○	○	○	○	○

Denken Sie, dass Sie die spezielle Befragungsmethode jeweils korrekt befolgt haben?
(mit Ausnahme der Beispielfrage zum Schwarzfahren)

bestimmt nicht	eher nicht	teils, teils	eher ja	ja, ganz bestimmt
○	○	○	○	○

Was ist Ihre persönliche Einschätzung:
Schützt die verwendete spezielle Befragungsmethode Ihre Antworten auf die heiklen Fragen zu 100%?

bestimmt nicht	eher nicht	teils, teils	eher ja	ja, ganz bestimmt
○	○	○	○	○

Für wie sinnvoll halten Sie den Einsatz dieser Befragungsmethode, um die Antworten der UmfrageteilnehmerInnen auf heikle Fragen zu schützen?

gar nicht sinnvoll	eher wenig sinnvoll	teils, teils	eher sinnvoll	sehr sinnvoll
○	○	○	○	○

Können Sie nachvollziehen, weshalb die verwendete spezielle Befragungsmethode Ihre Antworten zu 100% schützt?

nein, überhaupt nicht	eher nicht	teils, teils	eher ja	ja, ganz bestimmt
○	○	○	○	○

Appendix: additional tables

Prevalence estimates by technique

	copy	notes	drugs	partial	severe
Level					
DQ	17.8	9.1	3.4	2.9	1.5
	(1.2)	(0.9)	(0.6)	(0.6)	(0.5)
RRT	19.6	12.8	0.6	3.9	-0.6
	(1.2)	(1.1)	(1.0)	(1.2)	(1.1)
CM	27.4	14.7	9.6	7.8	3.1
	(2.0)	(1.9)	(1.9)	(2.1)	(2.1)
Difference					
RRT - DQ	1.8	3.7*	-2.7*	1.0	-2.1
	(1.7)	(1.4)	(1.1)	(1.3)	(1.2)
CM - DQ	9.6***	5.6**	6.3**	4.8*	1.5
	(2.3)	(2.1)	(2.0)	(2.2)	(2.1)
N	5841	5829	5809	4297	4291

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Prevalence estimates by implementation

	copy	notes	drugs	partial	severe
Level					
Direct questioning	17.8	9.1	3.4	2.9	1.5
	(1.2)	(0.9)	(0.6)	(0.6)	(0.5)
RRT random wheel	23.0	11.4	-0.9	1.2	0.7
	(2.1)	(2.0)	(1.7)	(2.0)	(2.0)
RRT pick a number	18.5	14.0	-1.6	2.9	-4.8
	(2.1)	(2.0)	(1.6)	(2.1)	(1.8)
RRT Benford	17.2	12.9	4.5	7.8	2.4
	(1.9)	(1.8)	(1.6)	(2.0)	(1.8)
CM unrelated quest~n	30.0	18.4	14.8	7.2	5.9
	(2.9)	(2.8)	(2.8)	(3.1)	(3.1)
CM pick a number	24.8	10.9	4.5	8.3	0.2
	(2.7)	(2.6)	(2.5)	(2.9)	(2.7)
Difference					
RRT rand. wheel - DQ	5.2*	2.3	-4.3*	-1.8	-0.8
	(2.5)	(2.2)	(1.8)	(2.1)	(2.1)
RRT pick number - DQ	0.7	4.9*	-5.0**	-0.1	-6.3***
	(2.4)	(2.2)	(1.7)	(2.2)	(1.9)
RRT Benford - DQ	-0.6	3.8	1.1	4.9*	0.8
	(2.3)	(2.0)	(1.7)	(2.1)	(1.9)
CM unr. quest. - DQ	12.2***	9.3**	11.4***	4.3	4.4
	(3.2)	(3.0)	(2.9)	(3.1)	(3.1)
CM pick number - DQ	7.0*	1.8	1.1	5.4	-1.3
	(3.0)	(2.7)	(2.5)	(3.0)	(2.8)
Observations	5841	5829	5809	4297	4291

Breakoff rates, response time and respondents' evaluation by implementation

	N	Breakoff	Time	Anonym.	NoRisk	Protect	Compreh.	Compl.
Direct questioning	1001	1.2	53.1	80.7	71.1			
		(0.3)	(1.5)	(1.3)	(1.4)			
RRT random wheel	1004	3.3	198.3	69.4	76.9	56.9	60.4	95.1
		(0.6)	(2.4)	(1.5)	(1.4)	(1.6)	(1.6)	(0.7)
RRT pick a number	1010	3.0	194.1	73.1	80.7	67.4	66.2	92.4
		(0.5)	(2.4)	(1.4)	(1.3)	(1.5)	(1.5)	(0.9)
RRT Benford	994	2.2	174.7	73.3	79.2	61.7	57.3	94.9
		(0.5)	(2.2)	(1.4)	(1.3)	(1.6)	(1.6)	(0.7)
CM unrel. question	1002	2.8	162.8	76.6	74.7	67.5	62.2	97.1
		(0.5)	(2.3)	(1.4)	(1.4)	(1.5)	(1.6)	(0.5)
CM pick a number	1001	3.2	198.4	76.6	80	75.0	65.6	95.7
		(0.6)	(2.5)	(1.4)	(1.3)	(1.4)	(1.5)	(0.7)

SE in parenthesis.
N: Number of assigned respondents
Breakoff: % who did not complete survey after reaching the sensitive questions
Time: Av. time (seconds) to answer the sensitive questions (highest 2.5 percentiles excluded)
Anonym.: % who trust in anonymity and privacy protection measures
NoRisk: % who think there is no disclosure risk
Protect: % who think their answers are protected thanks to RRT/CM
Compreh.: % who think they comprehend why RRT/CM protects their answers
Compl.: % who think they complied with RRT/CM procedure

Effects of respondents evaluation and respondents' protection on admitted cheating

Rand. resp. regression of cheating by type of cheating (all/exam/paper)

	logit_all	linear_~l	logit_ex	linear_ex	logit_pap	linear_~p
RRT	ref.	ref.	ref.	ref.	ref.	ref.
CM	0.32	6.30	0.37	9.83	1.30*	-0.47
	(0.54)	(6.27)	(0.52)	(7.57)	(0.52)	(9.60)
Trust in anonymity - ref.: RRT	0.30	1.73	0.33	3.43*		-1.70
	(0.17)	(1.42)	(0.17)	(1.70)		(2.16)
Trust in anonymity*CM	-0.13	-0.59	-0.17	-2.61		3.28
	(0.28)	(2.84)	(0.28)	(3.41)		(4.36)
No risk of disclosure - ref.: RRT	-0.11	-0.58	-0.12	-1.49		1.28
	(0.17)	(1.49)	(0.17)	(1.81)		(2.14)
No risk of disclosure*CM	0.081	1.88	-0.025	0.10		5.27
	(0.27)	(2.69)	(0.27)	(3.35)		(4.06)
Protects answers - ref.:RRT	-0.20	-0.24	-0.21	-0.63		0.68
	(0.16)	(1.45)	(0.17)	(1.76)		(2.13)
Protects answers*CM	-0.17	-3.08	-0.15	-3.85		-1.66
	(0.29)	(2.91)	(0.29)	(3.52)		(4.68)
Comprehension - ref.: RRT	0.16	1.04	0.20	1.94		-0.84
	(0.16)	(1.37)	(0.16)	(1.64)		(1.99)
Comprehension*CM	-0.093	-1.77	-0.079	-1.08		-3.09
	(0.25)	(2.65)	(0.26)	(3.20)		(4.25)
Compliance - ref.: RRT	-0.66**	-6.19*	-0.69**	-7.98*		-2.42
	(0.24)	(2.97)	(0.24)	(3.46)		(3.89)
Compliance*CM	0.39	1.63	0.39	1.91		1.17
	(0.53)	(6.17)	(0.51)	(7.32)		(9.42)
copy	ref.	ref.	ref.	ref.		
notes	-0.60***	-9.04***	-0.60***	-9.05***		
	(0.096)	(1.40)	(0.096)	(1.40)		
drugs	-1.95***	-18.4***	-1.95***	-18.4***		
	(0.24)	(1.42)	(0.24)	(1.42)		
partial	-1.56***	-16.9***				
	(0.20)	(1.51)				
severe	-3.08***	-22.0***				-5.15***
	(0.75)	(1.49)				(1.48)
Constant	-0.92***	25.0***	-0.92***	25.5***	-3.95***	6.77
	(0.27)	(3.26)	(0.26)	(3.79)	(0.46)	(4.24)
Observations	21085	21085	14127	14127	6958	6958

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001